

(19)日本国特許庁 (J P)

(12) 公開特許公報 (A)

(11)特許出願公開番号

特開平11-73406

(43)公開日 平成11年(1999)3月16日

(51)Int.Cl.⁸

G 0 6 F 15/18

識別記号

5 6 0

F I

G 0 6 F 15/18

5 6 0 A

審査請求 未請求 請求項の数11 O L (全 13 頁)

(21)出願番号 特願平10-169787

(22)出願日 平成10年(1998)6月17日

(31)優先権主張番号 08/883193

(32)優先日 1997年6月26日

(33)優先権主張国 米国 (US)

(71)出願人 596077259

ルーセント テクノロジーズ インコーポ
レイテッド

Lucent Technologies
Inc.

アメリカ合衆国 07974 ニュージャージ
ー、マレーヒル、マウンテン アベニュー
800-700

(72)発明者 クリストファー ジョン パージェス
アメリカ合衆国, 07728 ニュージャージ
ー、フリーホルド、アンドラ テラス
11

(74)代理人 弁護士 三侯 弘文

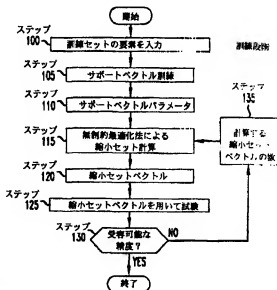
最終頁に続く

(54)【発明の名称】 サポートベクトル機械を使用する方法

(57)【要約】

【課題】 与えられたベクトルのセットを試験段階で用いるために高次元空間に写像するアルゴリズムを用いた機械の効率を改善する。

【解決手段】 サポートベクトル機械 (SVM) は、その判定面がサポートベクトルのセットによって、および、対応する重みのセットによって、パラメトライズされる万能学習機械である。本発明による SVM は、縮小セットベクトルを用いる。縮小セットベクトルの数は、セット内のベクトルの数より少ない。これらの縮小セットベクトルはセット内のベクトルとは異なり、同次2次核で用いられる固有値計算とは異なる最適化法に従って決定される。実施例では、パターン認識で用いるために、SVM は縮小セットベクトルを利用し、これにより、ユーザが選択するファクタだけこの SVM の効率を改善する。これらの縮小セットベクトルは、無制約最適化法に従って決定される。



【特許請求の範囲】

【請求項1】 入力データ信号を受け取るステップと、前記入力データ信号に作用可能なサポートベクトル機械を用いて出力信号を生成するステップとからなる、サポートベクトル機械を使用する方法において、前記サポートベクトル機械は縮小セットベクトルを利用し、

前記縮小セットベクトルは、同次2次核に用いられる固有値計算以外の最適化法を用いて訓練段階中であらかじめ決定されたものであることを特徴とする、サポートベクトル機械を使用する方法。

【請求項2】 前記訓練段階は、訓練セットの要素を受け取るステップと、 N_1 個のサポートベクトルからなるサポートベクトルセットを生成するステップと、 $m \leq N_1$ として、縮小セットベクトルの数 m を選択するステップと、無制約最適化法を用いて m 個の縮小セットベクトルを生成するステップとからなることを特徴とする請求項1に記載の方法。

【請求項3】 前記最適化法は無制約最適化法であることを特徴とする請求項1に記載の方法。

【請求項4】 前記入力データ信号は相異なるパターンを表し、前記出力信号は、該相異なるパターンの分類を表すことを特徴とする請求項1に記載の方法。

【請求項5】 前記訓練段階は、前記サポートベクトル機械を訓練してサポートベクトルの数 N_1 を決定するステップと、 $m \leq N_1$ として、無制約最適化法を用いて、 m 個の縮小セットベクトルを決定するステップとからなることを特徴とする請求項1に記載の方法。

【請求項6】 入力データ信号を提供するデータ入力要素と、前記入力データ信号に作用して少なくとも1つの出力信号を生成するサポートベクトル機械とからなる装置において、前記サポートベクトル機械は、同次2次核に用いられる固有値計算以外の最適化法を用いてあらかじめ決定された縮小セットベクトルを用いて前記入力データ信号に作用することを特徴とする、サポートベクトル機械を用いた装置。

【請求項7】 前記データ入力要素は、該データ入力要素に入力された複数の画像を表す入力データ信号を提供することを特徴とする請求項6に記載の装置。

【請求項8】 前記少なくとも1つの出力信号は、各画像の分類を表すことを特徴とする請求項7に記載の装置。

【請求項9】 前記縮小セットベクトルの数はサポートベクトルの数より少ないことを特徴とする請求項6に記載の装置。

【請求項10】 前記最適化法は無制約最適化法であることを特徴とする請求項6に記載の装置。

【請求項11】 前記縮小セットベクトルは、前記無制約最適化法を用いて前記サポートベクトル機械を訓練している間にあらかじめ決定されることを特徴とする請求項10に記載の装置。

【発明の詳細な説明】

【0001】

【発明の属する技術分野】 本発明は、万能学習機械に関し、特に、サポートベクトル機械に関する。

【0002】

【従来の技術】 サポートベクトル機械 (SVM (Support Vector Machine)) は、その判定面がサポートベクトルのセットによって、および、対応する重みのセットによって、パラメライズされる万能学習機械である。SVMはまた、核関数によっても特徴づけられる。核の選択は、その結果として得られるSVMが多項式クラシフィアであるか、2層ニューラルネットワークであるか、動径 (放射状) 基底関数 (RBF) 機械であるか、あるいはその他の学習機械であるかを決定する。SVMの判定規則は、対応する各関数およびサポートベクトルの関数である。

【0003】

【発明が解決しようとする課題】 一般に、SVMは、訓練段階および試験段階という2つの段階で動作する。訓練段階中、判定規則で用いるためのサポートベクトルのセットが生成される。試験段階中、特定の判定規則を用いて判定が行われる。残念ながら、この試験段階において、SVM判定規則の計算量は、サポートベクトルセット内のサポートベクトルの数 N_1 に比例する。

【0004】

【課題を解決するための手段】 本発明によれば、与えられたベクトルのセットを試験段階で用いるために高次元空間に写像するアルゴリズムを用いた機械の効率を改善する方法および装置が実現される。具体的には、本発明の原理によれば、縮小セットベクトルを用いる。縮小セットベクトルの数は、セット内のベクトルの数より少ない。これらの縮小セットベクトルはセット内のベクトルとは異なり、同次2次核で用いられる固有値計算とは異なる最適化法に従って決定される。

【0005】 本発明の実施例では、パターン認識で用いるために、SVMは縮小セットベクトルを利用し、これにより、ユーザが選択するファクタだけこのSVMの効率を改善する。これらの縮小セットベクトルは、無制約最適化法に従って決定される。

【0006】 本発明の特徴によれば、縮小セットベクトルの選択により、性能対計算量のトレードオフを直接制御することが可能となる。

【0007】 さらに、本発明の考え方はパターン認識に固有ではなく、サポートベクトルアルゴリズムが用いら

れるような任意の問題(例えば、回帰推定)に適用可能である。

【0008】

【発明の実施の形態】本発明の実施例について説明する前に、サポートベクトル機械について簡単な背景の知識を説明した後、本発明の考え方自体の説明を行う。本発明の考え方以外に、読者は、当業者に知られている核ベースの方法を一般的に表現するために用いられる数学的記法を知っていると仮定する。また、本発明の考え方は、パターン認識の場合の例に関して説明される。しかし、本発明の考え方は、サポートベクトルアルゴリズムが用いられるような任意の問題(例えば、回帰推定)に適用可能である。

【0009】以下の説明で、注意すべき点であるが、10個の数字のグレイレベル画像を含む2つの光学的文字認識(OCR)データセットからの試験データを用いる。一方のデータセットは、7,291個の訓練パターンおよび2,000個の試験パターンからなり、ここでは「郵便セット」という(例えば、L. Bottou, C. Cortes, H. Drucker, L. D. Jackel, Y. LeCun, U. A. Mueller, E. Saeckinger, P. Simard, and V. Vapnik, "Comparison of Classifier Methods: A Case Study in Handwritten Digit Recognition", Proceedings of the 12th *

$$y = \theta \left(\sum_{i=1}^{N_S} \alpha_i K(x, s_i) + b \right) \quad (1)$$

ただし、ベクトル x およびベクトル s_i は R^d の元であり、 α_i および b は実数であり、 θ は階段関数である。 R^d は d 次元ユークリッド空間であり、 R は実数である。 α_i 、ベクトル s_i 、 N_S および b はパラメータであり、ベクトル x は分類されるべきベクトルである。さまざまなクラシフィアに対する判定規則がこの形で書ける。例えば、 $K(x, s_i)$ は多項式クラシフィアを実現し、

【数2】

$$K = \exp(-\|x - s_i\|^2 / \sigma^2)$$

は動径基底関数機械を実現し、 $K = \tanh(y(x, s_i) + \delta)$ は2層ニューラルネットワークを実現する(例えば、V. Vapnik, "Estimation of Dependencies Based on Empirical Data", Springer Verlag, 1982, V. Vapnik, "The Nature of Statistical Learning Theory", Springer Verlag, 1995, Boser, B. E., Guyon, I. M., and Vapnik, V., "A training algorithm for optimal margin classifiers", Fifth Annual Workshop on Computational Learning Theory, Pittsburgh ACM 144-152, 1992, およびB. Schoelkopf, C. J. C. Burges, and V. Vapnik, "Extracting Support Data for a Given Task", Proceedings of the First International Conference on Knowledge Discovery and Data Mining, AA

* IAPR International Conference on Pattern Recognition, Vol. 2, IEEE Computer Society Press (米国カリフォルニア州ロサンゼルス), pp. 77-83, 1994, およびY. LeCun, B. Boser, J. S. Denker, D. Henderson, R. E. Howard, W. Hubbard, L. D. Jackel, "Backpropagation Applied to Handwritten ZIP Code Recognition", Neural Computation, 1, 1989, pp. 541-551, 参照)。他方のデータセットは、NIST Special Database 3およびNIST Test Data 1からの60,000個の訓練パターンおよび10,000個の試験パターンからなり、ここでは「NISTセット」という(例えば、R. A. Wilkinson, J. Geist, S. Janet, P. J. Grother, C. J. C. Burges, R. Greecy, R. Hammond, J. J. Hull, M. J. Larsen, T. P. Vogl and C. L. Wilson, "The First Census Optical Character Recognition System Conference, 米国商務省, NIST, August 1992, 参照)。「郵便セット」の画像は16×16ピクセルであり、「NISTセット」の画像は28×28ピクセルである。

【0010】[従来技術:サポートベクトル機械]判定規則が次の形をとるような2クラスクラシフィアを考える。

【数1】

AI Press (米国カリフォルニア州Menlo Park, 1995, 参照)。

【0011】サポートベクトルアルゴリズムは、その判定規則が式(1)の形をとるような任意の学習機械を訓練する原理的な方法である。要求される唯一の条件は、核 K が一般的な正値性制約を満たすことである(例えば、前掲の"The Nature of Statistical Learning Theory", および"A training algorithm for optimal margin classifiers", 参照)。他の方法とは異なり、SVM訓練プロセスは、パラメータセット全体 $\{\alpha_i, \text{ベクトル } s_i, N_S \text{ および } b\}$ を決定する。その結果得られるベクトル s_i ($i = 1, \dots, N_S$)は、訓練セットのサブセットであり、サポートベクトルと呼ばれる。

【0012】サポートベクトル機械はいくつかの優れた性質を有する。訓練手続きは、制約2次最適化問題を解くこととなり、従って、求められる解は、目的関数の一意的な大域的最小値であることが保証される。SVMは、構造的リスク最小化を直接に実現するために使用可能である。この場合、学習機械の容量は、汎化誤りの限界を最小にするように制御することができる(例えば、前掲の"The Nature of Statistical Learning Theory", および"Extracting Support Data for a Given Task", 参照)。サポートベクトル判定面は、実際には、高次元空間内の線形分離超平面である。同様に、SVM

は、回帰を構成するためにも使用可能であり、これはある高次元空間において線形である(例えば、前掲の“The Nature of Statistical Learning Theory”、参照)。

【0013】サポートベクトル学習機械は、光学的文字認識(OCR)(例えば、前掲の“The Nature of Statistical Learning Theory”、および“Extracting Support Data for a Given Task”、ならびに C. Cortes and V. Vapnik, “Support Vector Networks”, Machine Learning, Vol.20, pp.1-25, 1995、参照)、および対象認識のようなパターン認識問題に適用されて成功している。

【0014】図1は、従来技術のSVMの動作の流れ図である。この動作は、訓練段階および試験段階という2つの段階からなる。訓練段階では、ステップ52で、SVMは、クラスがあらかじめ割り当てられた訓練セットの要素を受け取る。ステップ54で、訓練セットからの入力データベクトルを多次元空間内へ変換する。ステップ56で、最適な多次元超平面に対するパラメータ(すなわち、サポートベクトルおよび対応する重み)が決定される。

【0015】図2に、訓練データ要素が2つのクラスに分離される例を示す。一方のクラスは円で表され、他方のクラスは四角で表されている。これは典型的な2クラスパターン認識問題のものである。例えば、「車」のパターンを「車でない」パターンから分離するように訓練されたSVMである。最適超平面は、2つのクラスのベクトルの間に極大マージンを有する線形判定関数である。すなわち、最適超平面は、訓練データを極大マージンで分離する一意的な判定面である。図2に示すように、最適超平面は、2つのクラスの間の分離が最大である領域によって定義される。図2で観察されるように、最適超平面を構成するには、訓練されたデータ要素のうち、この極大マージンを決定するサブセットを考慮すればよい。訓練要素のうち、最適超平面のパラメータを決定するこのサブセットは、サポートベクトルとして知られている。図2では、サポートベクトルは網掛けで示されている。

【0016】最適超平面は、高次元空間における写像されたサポートベクトルの線形結合で表される。SVMアルゴリズムは、ベクトルのセットに関する誤差が、すべてのサポートベクトルに重みを割り当てることによって最小化されることを保証する。これらの重みは、サポートベクトルによって判定面を計算する際にも用いられる。また、このアルゴリズムによれば、特定の問題に属する訓練データに関する誤り率を最小にするために、これらの重みを適応させることが可能となる。これらの重みは、SVMの訓練段階中に計算される。

【0017】このようにして、最適超平面を構成することは、訓練セットの要素および写像された空間内の内積を決定する関数によって決定される制約2次最適化計画問題になる。この最適化問題に対する解は、従来の中間

最適化法を用いて求められる。

【0018】一般に、最適超平面は、誤りなしで訓練データを分離することを必要とする。しかし、場合によっては、訓練データは誤りなしで分離することができないことがある。このような場合、SVMでは、最小数の誤りで訓練データを分離しようと試み、残りの要素を極大マージンで分離する。このような超平面は一般に、ソフトマージン超平面として知られている。

【0019】試験段階では、ステップ62で、SVMは、分類すべき試験セットの要素を受け取る。次に、SVMは、サポートベクトルを核のパラメータとして用いて、試験セットの入力データベクトルを多次元空間内へ写像することによって変換する(ステップ64)。写像関数は、SVMにあらかじめロードされている核の選択によって決定される。この写像は、1つのベクトルをとり、それを高次元特徴空間へ変換して、線形判定関数がこの高次元特徴空間に生成されるようにする。図1の流れ図は陰(implicit)の写像を示しているが、この写像は陽(explicit)に実行されることも可能である。ステップ66で、SVMは、各入力データベクトルの所属状態を示すように、判定面から分類信号を生成する。最終結果は、図2に示されるように、円の(+1)および四角の(-1)という出力分類信号の生成である。

【0020】残念ながら、式(1)の計算量は、サポートベクトルの数 N に比例する。サポートベクトルの数の期待値は $(1-E[P])$ で与えられる。ただし、 P は、与えられたSVMを1個の訓練サンプルで訓練した場合の、1つの試験ベクトルに対する誤りの確率であり、 $E[P]$ は、1個のサンプルのすべての選び方にわたる P の期待値である(例えば、前掲の“The Nature of Statistical Learning Theory”、参照)。従って、 N はおよそ1に比例することが予想される。実際のパターン認識問題では、この結果、同様の汎化性能を有する他のシステムよりも試験段階において大幅に悪い機械が得られる(例えば、前掲の“Comparison of Classifier Methods: A Case Study in Handwritten Digit Recognition”、および、Y. LeCun, L. Jackel, L. Bottou, A. Brunot, C. Cortes, J. Denker, H. Drucker, I. Guyon, U. Mueller, E. Saeckinger, P. Sinard, and V. Vapnik, “Comparison of Learning Algorithms for Handwritten Digit Recognition”, International Conference on Artificial Neural Networks, Ed. F. Fogelman, P. Gallinari, pp.53-60, 1995、参照)。

【0021】[縮小セットベクトル] これに対して、本発明の原理によれば、ずっと少数の縮小セットベクトルによりSVM判定規則を近似する方法および装置が実現される。縮小セットベクトルは以下の性質を有する。

【0022】・縮小セットベクトルは、サポートベクトルが完全なSVM判定規則に現れるのと同様にして、近似的なSVM判定規則に現れる。

7

・縮小セットベクトルは、サポートベクトルではない。
縮小セットベクトルは、サポートベクトルとは異なり、必ずしも分離マージン上にはなく、訓練サンプルでもない。

・縮小セットベクトルは、与えられた、訓練済みの SVM に対して計算される。

・縮小セットベクトルの数(従って、結果として得られる SVM の試験段階における速度)は事前に選択される。

・縮小セット法は、サポートベクトル法が用いられる場合であればどのような場合にも適用可能である(例えば、回帰推定)。

【0023】[縮小セット] 訓練データは、L の要素、ベクトル x であるとする。ただし、L (L は低次元(low dimensional)の意味)は、 d_L 次元ユークリッド空間 R^{d_L} として定義される。SVM は、陰写像

$$\Phi: x \mapsto \tilde{x}, \quad \tilde{x} \in H$$

を実行する。ただし、H (H は高次元(high dimensional)の意味)は R^H 、 $d_H \leq \infty$ である。以下では、H のベクトルにはバーを付けて示す。写像 Φ は、核 K の選択によって決定される。実際、Mercer の正値性制約(例えば、前掲の "The Nature of Statistical Learning Theory"、および "A training algorithm for optimal margin classifiers"、参照)を満たす任意の K に対して、*

$$\tilde{\Psi} \cdot \tilde{x}_i + b \geq k_0 - \xi_i, \quad y_i = +1 \quad (2)$$

$$\tilde{\Psi} \cdot \tilde{x}_i + b \leq k_1 + \xi_i, \quad y_i = -1 \quad (3)$$

ただし、 ξ_i は正のスラック変数であり、分離不能の場合(例えば、前掲の "Support Vector Networks"、参照)を扱うために導入したものである。分離可能の場合、SVM アルゴリズムは、H における正と負の例の間のマージンが最大になるような分離超平面を構成する。その後、

【数7】

$$\tilde{\Psi} \cdot \tilde{\Phi}(x) + b$$

が $(k_0 + k_1) / 2$ より大きい(小さい)かに応じて、試験ベクトル $x \in L$ にクラスラベル $\{+1, -1\}$ を割り当てる。サポートベクトル $s \in L$ は、式(2)または(3)のいずれかが等式になるような訓練サンプルとし*

$$\tilde{\Psi} \cdot \tilde{s} = \sum_{a=1}^{N_2} \alpha_a y_a \tilde{s}_a \cdot \tilde{s} = \sum_{a=1}^{N_2} \alpha_a y_a K(s_a, x) \quad (4)$$

【0025】しかし、本発明の考え方により、ここで、

【数10】

$$\Psi = \sum_{a=1}^{N_2} \gamma_a \Phi(s_a) \quad (6)$$

*【数4】

$$K(x_i, x_j) = \tilde{x}_i \cdot \tilde{x}_j$$

であるようなベア $\{\Phi, H\}$ が存在する。従って、H において、SVM 判定規則は単に(上記のように)、線形分離超平面となる。写像 Φ は通常、陽に計算されず、H の次元 d_H は通常大きい(例えば、同次写像 $K(x_i, x_j) = (x_i \cdot x_j)^p$ に対して、

【数5】

$$d_H = \binom{p + d_L - 1}{p}$$

である $(p + d_L - 1)$ 個のものから p 個のものを選ぶ場合の数。従って、4 次多項式で、 $d_L = 256$ の場合、 d_H は約 1.8 億となる)。

【0024】基本的な SVM パターン認識アルゴリズムは、2 クラス問題を解く(例えば、前掲の "Estimation of Dependencies Based on Empirical Data"、"The Nature of Statistical Learning Theory"、および "A training algorithm for optimal margin classifiers"、参照)。訓練データ $x \in L$ および対応するクラスラベル $y_i \in \{-1, 1\}$ が与えられた場合、SVM アルゴリズムは、ベクトル $x_i (i = 1, \dots, l)$ を 2 つのクラスに分ける判定面 $\Psi \in H$ を次のように構成する。

【数6】

30 ※で定義される。(サポートベクトルは、他の訓練データと区別するためにベクトル s と表す。)すると、 Ψ は次のように与えられる。

【数8】

$$\Psi = \sum_{a=1}^{N_2} \alpha_a y_a \Phi(s_a) \quad (7)$$

ただし、 $\alpha_a \geq 0$ は、訓練中に決定される重みであり、 $y_a \in \{-1, 1\}$ は、ベクトル s_a のクラスラベルであり、 N_2 はサポートベクトルの数である。こうして、試験ベクトル x を分類するためには、次式を計算する、

【数9】

が距離尺度

【数11】

$$\rho = \|\tilde{\Psi} - \tilde{\Psi}'\| \quad (8)$$

50 を最小にする(固定した N_2 に対して)ようなセット、

ベクトル $z_i \in L$ ($i = 1, \dots, N_i$) および対応する重み $y_i \in R$ を考える。

【0026】ここで、 $\{y_i, z_i\}$ ($i = 1, \dots, N_i$)

$$\Psi' \cdot \bar{x} = \sum_{i=1}^{N_i} \gamma_i \bar{x}_i \cdot \bar{x} = \sum_{i=1}^{N_i} \gamma_i h_i(x_i) \quad (8)$$

【0027】すると、目標は、結果として得られる汎化性能の損失が許容可能な範囲にとどまるような、最小の $N_i < N_i$ 、および対応する縮小セットを選択することである。明らかに、 $N_i = N_i$ とすることにより、 ρ を 0 にすることができる。しかし、 $N_i < N_i$ 、しかも $\rho = 0$ であるような自明でない場合が存在する（後述）。そのような場合、縮小セットにより、汎化性能の損失なしで、判定規則の計算量が低減される。各 N_i に対して、対応する縮小セットを計算する場合、 ρ は、 N_i の単調減少関数と見ることが可能であり、汎化性能もまた N_i の関数となる。本明細書では、汎化性能の N_i 依存性に関する経験的結果のみについて説明する。

【0028】写像 Φ について、以下のことに注意すべきである。 Φ の像は一般に線形空間にはならない。また、 Φ は一般に全射にはならず、一対一でない可能性がある（例えば、 K が偶数次の同次多項式の場合）。さらに、 Φ は、 L 内の線形従属ベクトルを M 内の線形独立ベクトルに写像することがあり得る（例えば、 K が非同次多項式の場合）。 K が同次多項式の場合であっても、一般に、ベクトル z_i をスケールすることによって係数 γ_i を 1 にスケールすることはできない（例えば、 K が偶数次の同次式である場合、 γ_i は $\{+1, -1\}$ にスケールすることは可能であるが、必ずしも 1 にスケールすることはできない）。

【0029】〔厳密解〕このセクションでは、 ρ の最小値を解析的に計算する問題を考える。まず、簡単ではあるが自明ではない場合について説明する。

【0030】〔同次 2 次多項式〕同次 2 次多項式の場合、規格化を 1 に選ぶ。

$$K(x_i, x_i) = (x_i \cdot x_i)^d \quad (9)$$

【0031】説明を簡単にするため、1 次近似 $N_i = 1$ を計算する。対称テンソル

$$S_{\mu\nu} \equiv \sum_{i=1}^{N_i} \alpha_i y_i x_{i\mu} x_{i\nu} \quad (10)$$

を導入する。

$$\rho^2 = S_{\mu\nu} S^{\mu\nu} - \sum_{i=1}^{N_i} \gamma_i^2 \|z_i\|^4 \quad (11)$$

ただし、固有ベクトルは、固有値の絶対値の大きさの順に並べるものとする。なお、 $\text{trace}(S)$ は、 S の固有値の平方の和であるので、 $N_i = d$ （データの次元）と選択することにより、近似は厳密（すなわち $\rho =$

* N_i ）を縮小セットという。試験点ベクトル x を分類するには、式（5）の展開を次の近似で置き換える。

【数12】

* 【数14】

$$\rho = \|\bar{\Psi} - \gamma \bar{z}\|$$

は、次式を満たす $\{\gamma, \text{ベクトル } z\}$ に対して最小になることが分かる。

【数15】

$$S_{\mu\nu} z_\nu = \gamma z_\mu^2 \quad (11)$$

（繰り返す添字については和をとる）。 $\{\gamma, \text{ベクトル } z\}$ をこのように選ぶと、 ρ^2 は次のようになる。

【数16】

$$\rho^2 = S_{\mu\nu} S^{\mu\nu} - \gamma^2 z^4 \quad (12)$$

【0032】従って、 $\{\gamma, \text{ベクトル } z\}$ を、ベクトル z が、 S の固有値 $\lambda = \gamma z^4$ が最大絶対値を有するような固有ベクトルとなるように選択するときに、 ρ の最大降下が達成される。なお、 γ は、 $\gamma = \text{sign}(\lambda)$ となるように選択することが可能であり、ベクトル z は $z^4 = |\lambda|$ となるようにスケールすることが可能である。

【0033】オーダー N_i に拡張すると、同様にして、式

30 【数17】

$$\rho = \left\| \bar{\Psi} - \sum_{i=1}^{N_i} \gamma_i \bar{z}_i \right\| \quad (13)$$

を最小にするセット $\{\gamma_i, \text{ベクトル } z_i\}$ におけるベクトル z_i は、それぞれ固有値が

【数18】

$$\gamma_i \|z_i\|^2$$

である S の固有ベクトルであることが示される。これにより次式が成り立ち、 ρ の降下は、ベクトル z_i を S のはじめの N_i 個の固有ベクトルに選択した場合に最大となる。

【数19】

0) になる。サポートベクトルの数 N_i は、 d より大きいことが多いため、このことは、汎化性能の損失なしに、縮小セットのサイズはサポートベクトルの数より小さくなりうることを示している。

11

【0034】一般の場合、縮小セットを計算するために、 ρ は、すべての $\{y_i, \text{ベクトル } z_i\}$ ($i = 1, \dots, N_i$) にわたって同時に最小にならなければならない。次のような反復法を考えると便利である。すなわち、第1ステップでは、 $\{y_i, \text{ベクトル } z_i\}$ ($j < i$) を固定して、 $\{y_i, \text{ベクトル } z_i\}$ を計算する。2次多項式の場合、この反復法によって生成される最小値の列が、問題全体に対する最小値も生成する。この結果は、2次多項式に特有であり、ベクトル z_i が直交する (あるいはそのように選択することができる) という事実の結果である。

【0035】以下の表1に、試験セットに関して誤りの数 E を達成するために必要な縮小セットサイズ N_z を示す。ここで、郵便セットに関して訓練された2次多項式 SVM の場合、 E_z は、サポートベクトルの完全セットを用いて求められる誤りの数 E_s とは、高々1個の誤りしか異なる。明らかに、2次の場合、縮小セットは、精度をほとんど失うことなく、計算量を大幅に減らすことができる。また、多くの数字では、サポートベクトルの数は $d_i = 256$ より大きい。これは、精度を全く失わずに高速化が可能であることを示す。

【表1】

数字	サポートベクトル		縮小セット	
	N_s	E_s	N_z	E_z
0	292	15	10	16
1	95	9	6	9
2	415	28	22	29
3	403	26	14	27
4	375	35	14	34
5	421	26	18	27
6	261	13	12	14
7	228	18	10	19
8	446	33	24	33
9	330	20	20	21

30

$$S_{\mu_1 \mu_2 \dots \mu_n} z_{1\mu_1} z_{2\mu_2} \dots z_{n\mu_n} = \gamma_1 \|z_1\|^{2n-2} z_{2\mu_1} \quad (17)$$

ただし、

※ ※ 【数22】

$$S_{\mu_1 \mu_2 \dots \mu_n} \equiv \sum_{m=1}^{N_E} \alpha_m y_m \varepsilon_{m\mu_1} \varepsilon_{m\mu_2} \dots \varepsilon_{m\mu_n} \quad (16)$$

である。

【0037】この場合、 γ に関して ρ を変化させて新しい条件は得られない。1次の解 $\{y_i, \text{ベクトル } z_i\}$ ★ 【数23】

$$\rho^2 = S_{\mu_1 \mu_2 \dots \mu_n} S_{\mu_1 \mu_2 \dots \mu_n} - \gamma_1^2 \|z_1\|^{2n} \quad (17)$$

【0038】そこで、次のような定義をすることができる。 ☆ 【数24】

$$\tilde{S}_{\mu_1 \mu_2 \dots \mu_n} \equiv S_{\mu_1 \mu_2 \dots \mu_n} - \gamma_1^2 z_{1\mu_1} z_{1\mu_2} \dots z_{1\mu_n} \quad (18)$$

12

* 【0036】【一般の核】縮小セット法を任意のサポートベクトル機械に適用するには、上記の解析を一般の核に拡張しなければならない。例えば、同次多項式 $K(x_1, x_2) = N(x_1 \cdot x_2)^n$ の場合、反復法の最初のベア $\{y_i, \text{ベクトル } z_i\}$ を求めるために

【数20】

$$\frac{\partial \rho}{\partial z_{1a_1}} = 0$$

とおくと、式(11)に類似の次式が得られる。

【数21】

13

これにより、2次の解ベクトル z_i に対する反復方程式が、式(15)で S 、ベクトル z_i 、および y_i をそれぞれ $\sim S$ 、ベクトル z_i 、および y_i で置き換えた形をとる。

(なお、2より高い次数の多項式では、ベクトル z_i は一般に直交しない。)しかし、これらは反復解のみであり、さらに、すべての $\{y_i, \text{ベクトル } z_i\}$ が同時に変化することを許容した場合の連立方程式を解く必要がある。さらに、これらの方程式は複数の解を有し、そのほとんどは ρ に関する極小値に対応する。さらに、別の K を選択することにより、他の固定点方程式が得られる。*10

$$\frac{\partial F}{\partial x_k} = - \sum_{m=1}^{N_2} \alpha_m y_m K(s_m \cdot x_k) + \sum_{j=1}^{N_2} \gamma_j K(z_j \cdot x_k) \quad (17)$$

$$\frac{\partial F}{\partial z_{jk}} = - \sum_{m=1}^{N_2} \gamma_k \alpha_m y_m K'(s_m \cdot x_k) z_{jm} + \sum_{j=1}^{N_2} \gamma_j \gamma_k K'(z_j \cdot x_k) z_{jk} \quad (18)$$

【0040】従って、本発明の原理によれば、(おそくは局所的な)最小は、無制約最適化法を用いて求めることができる。

【0041】【アルゴリズム】まず、所望の近似次数 N を選択する。 $X_i = \{y_i, z_i\}$ とする。2段階法を用いる。第1段階(後述)で、すべてのベクトル z_i ($j < i$)を固定したまま、 X_i を反復的に計算する。

【0042】第2段階(後述)で、すべての X_i が変動することを許容する。

【0043】注意すべき点であるが、式(20)における勾配は、 y_i が0である場合、0である。この事実とは、単大な数値的不安定性につながる可能性がある。こ

$$\Gamma_j = \gamma_j$$

$$\Delta_j = \sum_{m=1}^{N_2} \alpha_m y_m h(s_m \cdot z_j) \quad (21)$$

$$Z_{jk} = K(z_j, x_k) \quad (22)$$

【0044】 Z は正定値かつ対称であるため、周知のコレスキ分解を用いて効率的に逆行列を求めることができる。

【0045】こうして、アルゴリズムの第1段階は以下のように進行する。

[1] $y_i = +1$ または -1 をランダムに選び、ベクトル z_i をランダム値に設定する。

[2] ベクトル z_i を変化させて F を最小化する。

[3] ベクトル z_i を固定したまま、 F をさらに最大に低下させる y_i を計算する。

[4] ベクトル z_i, y_i をともに変化させてさらに F を低下させる。

[5] 最良の解を保持してステップ[1]～[4]を T 回反復する。

[6] ベクトル z_i, y_i を固定し、 $y_i = +1$ または -1 をランダムに選び、ベクトル z_i をランダム値に設定する。

14

* 式(15)の解は反復(すなわち、任意のベクトル z から始めて、式(15)を用いて新たなベクトル z を計算し、これを繰り返す)によって求められるが、次のセクションで説明する方法はさらに柔軟で強力である。

【0039】【無制約最適化法】核 K の1次導関数が定義されていると仮定すると、未知数 $\{y_i, \text{ベクトル } z_i\}$ に関する目的関数 $F = \rho^2/2$ の勾配を計算することができる。例えば、 $K(s_i, s_j)$ がスカラー $s_i \cdot s_j$ の関数であると仮定すると、次式のようになる。

【数25】

※の問題を回避するために、第1段階は、単純な「レベル交差」定理に基づく。そのアルゴリズムは以下のとおりである。まず、 y_i を $+1$ または -1 に初期化し、ベクトル z_i をランダム値で初期化する。次に、 y_i を固定したまま、ベクトル z_i を変化させる。次に、ベクトル z_i, X_i ($j < i$)を固定した場合の、 y_i の最適値を解析的に計算する。次に、ベクトル z_i 、および y_i の両方に関して同時に F を最小化する。最後に、すべての $j \leq i$ に対して最適な y_i を解析的に計算する。これは、 $\Gamma = Z^{-1} \Delta$ によって与えられる。ただし、 δ, Γ および Z は次のように与えられる(式(19)参照)。

【数26】

$$\Gamma_j = \gamma_j$$

$$\Delta_j = \sum_{m=1}^{N_2} \alpha_m y_m h(s_m \cdot z_j)$$

$$Z_{jk} = K(z_j, x_k)$$

[7] ベクトル z_i を変化させて F を最小化する。

[8] ベクトル z_i 、およびベクトル z_i, y_i を固定し、 F をさらに最大に低下させる最適な y_i を計算する。

[9] $\{y_i, \text{ベクトル } z_i, y_i\}$ をともに変化させてさらに F を低下させる。

[10] 最良の解を保持してステップ[6]～[9]を T 回反復する。

[11] 最後に、ベクトル z_i 、ベクトル z_i を固定し、(上記の式(21)～(23)に示されるように)さらに F を低下させる最適な y_i, y_i を計算する。

【0046】次に、この手続きを $\{\text{ベクトル } z_i, y_i\}$ 、 $\{\text{ベクトル } z_i, y_i\}$ など、

【数27】

$$\{z_{N+1}, \gamma_{N+1}\}$$

50 まで反復する。

【0047】 γ_i が0に近づかないようにすることによって数値的不安定性は回避される。上記のアルゴリズムにより、これは自動的に保証される。第1ステップで、 γ_i を固定したままベクトル z_i を変化させた結果、1の関数 F が減少した場合、次に γ_i を変化させるときに、 γ_i は0を通ることはできない。その理由は、0を通るとすると(その場合[ベクトル z_i 、 γ_i]の F への寄与は0となるので) F が増大してしまうからである。

【0048】なお、与えられた[ベクトル z_i 、 γ_i]のペアの各計算は、第1段階で、ベクトル X_i に対する相異なる初期値で数回(T回)反復される。Tは、求められたFにおける相異なる最小値の個数Mから経験的に決定される。上記のデータセットでは、Mは通常2または3であり、Tは $T=10$ と選ばれた。

【0049】第2段階では、第1段階で求められたすべてのベクトル X_i が単一のベクトルへと接続され、すべてのパラメータの変動を許容して、再び無制約最小化プロセスが適用される。注意すべき点であるが、第2段階の結果、目的関数Fがさらに約2分の1に減少することが多い。

【0050】本発明の原理に従って、以下の1次無制約最適化法を両方の段階で用いた。探索方向は、共役勾配法を用いて求められる。探索方向に沿って、ブラケット点 x_1 、 x_2 および x_3 を、 $F(x_1) > F(x_2) < F(x_3)$ となるように求める。次に、このブラケットを平衡化する(平衡化法については、例えば、W. H. Press, S. A. Teukolsky, W. T. Vetterling and B. P. Flannery, "Numerical Recipes in C", Second Edition, Cambridge University Press, 1992、参照)。これらの3点を通る2次当てはめ曲線の最小値を、次の反復で選択される開始点として用いる。共役勾配プロセスは、所定の反復数の後に再開され、全体のプロセスは、Fの減少率があるしきい値を下回ったときに終了する。注意すべき点であるが、この一般的なアプローチは、上記の2次多項式核の場合に適用した場合、解析的アプローチと同じ結果を与えた。

【0051】[実験]上記のアプローチを、郵便セットに対して最良の性能を有するSVMに適用した。このSVMは次数3の非同次多項式機械(これについては、例えば、前掲の"The Nature of Statistical Learning Theory"、参照)であった。近似の次数Nは、各2クラスクラシファイアに対して試験段階で10倍の高速化がなされるように選択した。結果を表2(下記)に示す。縮小セット法は、精度の損失はほとんどなしで、この高速化を達成した。10個のクラシファイアをまとめて1つの10クラスクラシファイア(これについては、例えば、前掲の"The Nature of Statistical Learning Theory"、および"Support Vector Networks"、参照)として用いると、完全サポートセット(サポートセット全体)を用いた場合には、4.2%のエラーであるのに対して、

縮小セットを用いた場合には、4.3%のエラーであった。なお、組み合わせた場合、縮小セットでは6倍の高速化が得られない。その理由は、相異なる2クラスクラシファイアがいくつかの共通のサポートベクトルを有し、キャッシングの可能性があるためである。これらの方法をさらに大きい問題に拡張することができかどうかという問題を解決するため、NISTセットの場合に、数字0を他のすべての数字から分離する2クラスクラシファイアに対して研究を繰り返した(60,000回の訓練、10,000個の試験パターン)。このクラシファイアも、完全サポートセットを用いて、最良の精度を与えるもの(次数4の多項式)を選んだ。1,2,7,3個のサポートベクトルの完全セットでは19個の試験エラーを生じたが、サイズ127の縮小セットでは20個の試験エラーであった。

【0052】

【表2】

数字	サポートベクトル		縮小セット	
	N_S	E_S	N_Z	E_Z
0	272	13	27	13
1	109	9	11	10
2	380	26	38	26
3	418	20	42	20
4	392	34	39	32
5	397	21	40	22
6	257	11	26	11
7	214	14	21	13
8	463	26	46	28
9	387	13	39	13
計	3289	187	329	188

【0053】(なお、試験は、完全な10桁のNISTに対しても行われ、10%の精度損失で50倍の高速化がなされた。C. J. C. Burges, B. Schoelkopf, "Improving the Accuracy and Speed of Support Vector Machines", in press, NIPS '96、参照。)

【0054】

[実施例]図3に、SVMの訓練段階で用いられる、本発明の原理を実現する例示的な流れ図を示す。ステップ100で、入力訓練データがSVM(図示せず)に入力される。ステップ105で、SVMがこの入力データに対して訓練され、ステップ110で、SVMはサポートベクトルのセットを生成する。ステップ115で、縮小セットベクトルの数が選択される。ステップ115で、無制約最適化法(前述)を用い、ステップ120で縮小セットベクトルを生成する。ステップ125で、この縮小セットベクトルを用いて、サンプルデータのセット(図示せず)を試験する。ステップ130で、この試験の結果を評価する。試験結果が(例えば速度および精度に関して)受容可能な場合、この縮小セットベクトルが

以後利用される。試験結果が受容可能でない場合、縮小セットベクトルを決定するプロセスを再び実行する。
(後者の場合、注意すべき点であるが、(例えば速度あるいは精度に関する)試験結果は、縮小セットベクトルの数をさらに少なくすることを示唆する可能性もある。)

【0055】縮小セットベクトルが決定されると、SVMで利用可能となる。この縮小セットベクトルを試験段階で使用方法を図4に示す。ステップ215で、試験セットからの入力データベクトルがSVMに送られる。ステップ220で、SVMは、縮小セットベクトルを核のパラメータとして用いて、試験セットの入力データベクトルを多次元空間に写像することにより変換する。ステップ225で、SVMは、判定面から、各入力データベクトルの帰属状態を示す分類信号を生成する。

【0056】上記のように、m個の縮小セットベクトルが縮小セット内にある。これらの縮小セットベクトルは、図3に示した上記の訓練段階で決定される。速度および精度のデータが、m個より少ない縮小セットベクトルを使用することも可能であることを示唆する場合、別のアプローチを用いて、新たなさらに小さい縮小セットベクトルのセットを再計算する必要を回避することが可能である。特に、 $x < m$ として、x個の縮小セットベクトルは、m個の縮小セットベクトルのセットから選択される。この場合、いくつ(x)の縮小セットベクトルを使用するかの決定は、例えば、訓練段階で生成された速度および精度のデータを用いて経験的に行われる。しかし、これらの縮小セットベクトルの値を再計算する必要はない。

【0057】パターン認識の場合の、本発明の考え方の実施例を図5に示す。パターン認識システム100は、プロセッサ105および認識器110からなり、認識器110は、データ入力要素115、およびSVM120からなる。本発明の考え方以外には、図5の要素は周知であるため、詳細には説明しない。例えば、データ入力要素115は、分類するための入力データをSVM120へ送る。データ入力要素115の一例はスキャナである。この場合、入力データは画像のピクセル表現(図示せず)である。SVM120は、本発明の原理に従って縮小セットベクトルを用いて入力データに作用する。動作(試験)中、SVM120は、入力データの分類を表す数値結果を、後続の処理のためにプロセッサ105に

送る。プロセッサ105は、例えば、メモリを伴うマイクロプロセッサのような蓄積プログラム制御プロセッサである。プロセッサ105は、さらに、例えば自動預払機(ATM)などにおける認識器110の出力信号を処理する。

【0058】図5のシステムは2つのモード、すなわち、訓練モードおよび動作(試験)モードで動作する。訓練モードの例は、図3に示される上記の方法である。試験モードの例は、図4に示される上記の方法である。

【0059】以上、本発明について説明したが、当業者には認識されるように、本発明の技術的範囲内でさまざまな変形例を考えることができる。例えば、本発明の考え方は、サポートベクトル機械以外の、核に基づく方法にも適用可能であり、例えば、回帰推定、密度評価などにも使用可能であるが、これらに限定されるものではない。

【0060】

【発明の効果】以上述べたごとく、本発明によれば、与えられたベクトルのセットを試験段階で用いるために高次元空間に写像するアルゴリズムを用いた機械の効率を改善する方法および装置が実現される。本発明の特徴によれば、縮小セットベクトルの選択により、性能対計算量のトレードオフを直接制御することが可能となる。さらに、本発明の考え方はパターン認識に固有ではなく、サポートベクトルアルゴリズムが用いられるような任意の問題(例えば、回帰推定)に適用可能である。

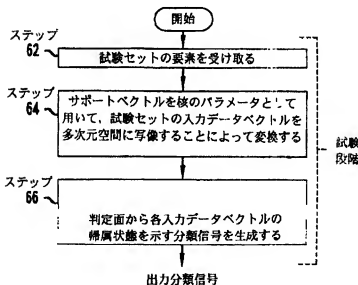
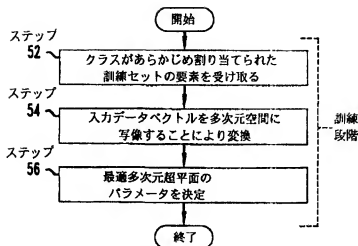
【図面の簡単な説明】

- 【図1】従来技術のSVMの動作の流れ図である。
- 【図2】代表サポートベクトルにより訓練データを2つのクラスに分離する一般的な図である。
- 【図3】本発明の原理に従ってSVMシステムを訓練する例示的な方法の図である。
- 【図4】本発明の原理に従ってSVMシステムを動作させる例示的な方法の図である。
- 【図5】本発明の原理を実現する認識システムの一部のブロック図である。
- 【符号の説明】

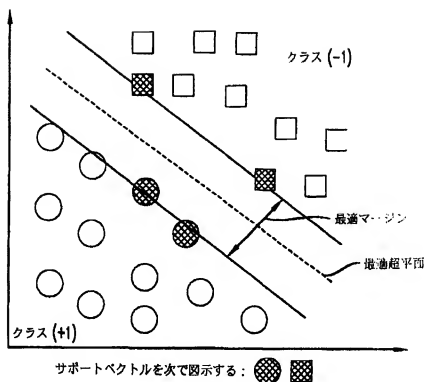
- 100 パターン認識システム
- 105 プロセッサ
- 110 認識器
- 115 データ入力要素
- 120 SVM

【図1】

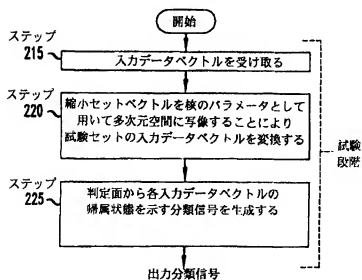
(従来技術)



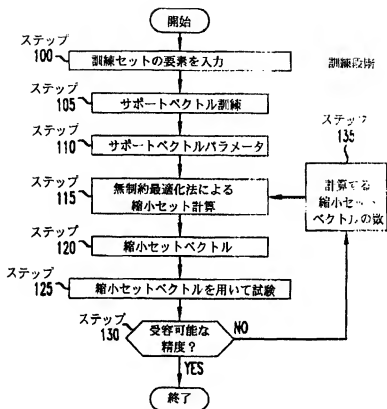
【図2】



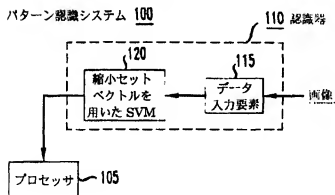
【図4】



【図3】



【図5】



フロントページの続き

(71)出願人 596077259

600 Mountain Avenue,
Murray Hill, New Je
rsey 07974-0636 U. S. A.